DOCUMENT RESUME

ED 246 090 TM 840 362

AUTHOR Wilson, Barry J.; Reichmuth, Monica
TITLE Early Screening Programs: When Is Predictive Accuracy Sufficient?
PUB DATE Apr 84
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans, LA, April 23-27, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Information Analyses (070) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Handicap Identification; *High Risk Students; Predictive Measurement; *Predictive Validity; *Preschool Children; Preschool Education; Reliability; Screening Tests; Student Placement
IDENTIFIERS Accuracy

ABSTRACT

Early screening programs to determine learning problems in preschool children are evaluated. A review of early identification literature indicates considerable difficulty in describing screening results and determining when a screening program accurately identifies those students believed to be "at risk." The authors propose several criteria to assist the determination and communication of screening results. These criteria are applied to a screening program for a metropolitan school district, the Omaha Parochial Schools. They provide a context for judging the accuracy of predictions. However, it is argued that accuracy of prediction is of little value, and cannot be judged sufficient, unless the predictions made about children afford better educational opportunities for these students. (Author/DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED246090

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

EARLY SCREENING PROGRAMS: WHEN IS PREDICTIVE ACCURACY SUFFICIENT?

Barry J. Wilson
University of Northern Iowa

Sister Monica Reichmuth
College of St. Mary

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

B J Wilson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

American Educational Research Association
Annual Meeting
New Orleans
April 1984

TM 840 362

EARLY SCREENING PROGRAMS: WHEN IS PREDICTIVE ACCURACY SUFFICIENT?

Barry J. Wilson
Monica Reichmuth

ABSTRACT

Review of early identification literature indicates considerable problems in describing screening results and determining when a screening program accurately identifies those students believed to be "at-risk". Several criteria are proposed to assist the determination and communication of screening results and these criteria are applied to a screening program for a metropolitan school district. These criteria provide a context for judging accuracy of prediction but it is argued that accuracy of prediction is of little value and cannot be judged sufficient unless the predictions made about children afford better educational opportunities for these students.

Attempts to identify preschool children who are likely to encounter difficulties in school learning are analagous to the age-old problem of deciding which melon in the melon patch is "just right" and which are likely to be "not ready" or defective. Rumors abound that there are semi-scientific methods of identifying good melons and good students. Most melon-thumpers will admit that they are occasionally wrong and the same is true of those who attempt to sort children. Essential characteristics are often covert and decisions are typically made on the basis of inference. Knowledge of growing conditions can be helpful in either situation and some "problem" melons and children are more detectable than others. The decisions to be made about children are more complex, difficult, and of far greater import to society. Thus our attempts to develop and refine decision strategies for children that go beyond folklore are of major significance to educators and to children. The purpose of the present paper is to suggest ways that we may improve our methods of assessing and communicating about the accuracy of prediction and to emphasize the examination of our prediction efforts in the context of decision strategies.

Preschool and kindergarten screening programs have become increasingly popular among school districts (Mercer, 1983). These programs have been justified by the idea that early identification of learning problems may allow appropriate intervention to be made so that these problems can be prevented or alleviated (Satz & Fletcher, 1979). While this goal is generally viewed as commendable, a number of investigators have pointed out the dangers of making such predictions (Keogh & Becker, 1973; Lichenstein, 1982). These include the problems of predictive accuracy and the relevance of screening information for appropriate intervention. Furthermore, screening measures are often used to make decisions about children without any systematic evaluation of their effectiveness. There is also a danger that students identified as "at-risk" may be subjected to inappropriate expectations by teachers and parents as a result of labeling.

The problem of predictive accuracy is particularly accute. Recent reviewers have pointed out that many studies have reached conclusions about predictive effectiveness on the basis of incomplete analysis of data (Mercer, Algozzine, & Trifiletti, 1979) and have often failed to consider base rate information as a criteria for predictive effectiveness (Satz & Fletcher, 1979). It is clear that if screening and selection procedures are not accurate in detecting students who are at risk, those procedures are unlikely to serve a useful purpose. The determination of acceptable accuracy is not, however, a simple task and review of the current literature suggests considerable variation concerning

acceptable indices of predictive accuracy. For example, Glazzard (1980) concluded that a reading readiness test and a kindergarten rating scale were acceptiable predictors of reading achievement in first through fourth grade. The correlations that were reported were indeed statistically significant from zero correlation but virtually identical to those reported in an earlier study by Eaves, Kendall, and Crichton (1974) who were much more pessimistic with respect to the instruments used in their study. Similar contrasts in conclusions characterize studies that report predictive efficiency in terms of hit rate. These contrasts have prompted Keogh (1977) to describe conclusions made about predictive accuracy as "selective perception".

Definition of high-risk. Review of the literature on predictive effectiveness in the identification of at risk learners suggests considerable variation, even confusion, with respect to methodology, assessment of findings, decision contexts, and intervention strategies. A good portion of this confused state can be attributed to the difficulty in defining target populations. What precisely are the "risks" that face the preschool child? Is it the risk of becoming below average in one or more school subjects, the risk of being seen as inattentive and/or disruptive in the classroom, the risk of being described as learning disabled, mentally deficient, or emotionally disturbed and thus requiring expensive intervention at a later date? If we fail to be clear about what state we are attempting to predict, we cannot be clear about what we mean when we say that a preschool child is at-risk and several authors have correctly questioned the utility of the term "at-risk". While this problem is not the major focus of this paper, these considerations have a major impact on the problem of prediction.

Prediction Methodologies. The traditional concept of predictive validity has resulted in numberless studies in the early identification literature that have addressed the question of predictive validity as expressed by a validity coefficient. While the predictive validity coefficient can have some utility in the selection of a test or procedure, there are also some limitations that may be clarified by returning to the melon patch analogy. One could be informed by a sage melon farmer that the Three-Thump test has a correlation of .68 with taste ratings by a panel of melon gourmets. The information is less than informative when the consumer is interested only in the odds of getting a good melon. To be sure, the history of psychological testing has provided some notable and potentially useful help in making this determination (Taylor and Russell, 1939), but validity coefficients in themselves focus attention on the congruence of test or rating scores without clarifying the degree that

selection decisions made about individuals are valid. The purpose of most screening programs has been to select some students for what is hoped to be appropriate educational intervention and it is the selection and decision process that needs to be validated. Studies that report only validity coefficients do not directly address selection and decision variables although the fundamental question of test validity may well be one of several important considerations.

Prediction and Hit Rate. Assessment of predictive efficiency has also been reported using other prediction models. Currently, the most prevalent of these is the prediction-performance matrix. Figure 1 displays the typical 2 x 2 matrix:

	low achievers	OK achievers
At-Risk	true positive	false negative
Ready	false positive	true negative

This model was originally proposed by Meehl and Rosen (1955) and has influenced the work of a number of researchers in the early identification area (Satz and Fletcher, 1980; Barnes, 1982). The model has also inflenced selection procedures and research in the areas of clinical and industrial psychology. The model directs attention to decision efficiency. This model is deceptively simple, however, and can be interpreted (as well as misinterpreted) in a number of ways. A numerical example using the data by Reichmuth (1983) will help to clarify the use of the model.

Reichmuth evaluated decisions made about 113 students in the Omaha parochial school system using locally developed screening procedures. Results using this model report results in terms of a hit rate where hit rate is defined as the percentage of children correctly classified as at-risk (positives) or not at-risk (negatives). The number of true positives (at-risk who later were classified as problem

learners) and true negatives (not at-risk who were later classified as good learners) divided by the total number screened yields an overall hit rate. Hit rates for false positives and false negatives are also typically reported and these can be computed horizontally, vertically, or as a percentage of the total sample. As indicated in Table 1, the overall hit rate was 8 + 83 / 113 = 80.5 per cent. The percentage of false positives was 13 of 21 or 81.5 per cent

---

Insert Table 1 About Here

---

using the row total as the denominator (horizontal method) or 13 of 96 or 13.5 per cent using the column total as the denominator (vertical method). This can also be computed as 13 of 113 using the total sample although this type of computation is less frequently encountered in the early identification literature. The fact that different methods of computation lead to different percentages has caused some confusion in the literature (Satz & Fletcher,1979).

Mercer (1983) has pointed out that while each method yields different information, both horizontal and vertical methods can be important in the interpretation of results. However, researchers have often been selective in reporting this information and in some cases may present only a portion of the analysis.

Factors Affecting Hit Rate. The relative simplicity of the model is further clouded by the fact that predictive efficiency is significantly affected by base rate and selection ratio as well as test or procedure validity (Wiggins, 1973; Anastasi,1982). The proportion of problem learners in a specific school age population (base rate) may fluctuate dramatically depending on district characteristics and criteria employed. Meehl and Rosen (1955) were first to point out the effects of base rate on predictive efficiency and also noted that validity data from one population with a given base rate cannot be generalized to populations with markedly different base rates.

Base rate is defined as the incidence of a particular condition within the population. The base rates of certain conditions can be determined with some degree of objectivity (e.g. incidence of deafness in a school age population). In the case of learning problems, however, determination of base rate is a much more subjective decision for reasons indicated earlier. If, for example, one were to attempt to predict students in third grade who are "learning disabled", a base rate of 3 to 5 per cent might well be appropriate.

Attempting to predict students in third grade who have "reading problems" might result in a base rate of 10 to 20 per cent or more. Many studies report base rates in excess of these figures including some that have identified more than 50 per cent of the population as "low achieving" (Ferenden & Jacobsen,1970).

The Base Rate Criterion. The base rate can yield a rather severe criteria for evaluating predictive effectiveness when the incidence or base rate is relatively low. Thus, if a base rate is determined to be 10 per cent, an overall hit rate of 90 per cent could be obtained by declaring everyone to be "not at-risk". Thus a screening procedure that successfully predicted 85 per cent of the students would fail to meet the base rate criterion by 5 per cent. As Lindsay (1982) has noted, few studies have examined the question of whether screening procedures have improved on the prediction derived from base rate. Failure to meet this criteria does not in itself suggest that a screening procedure is of no value since prediction from base rate would result in no one being selected for intervention. When intervention results in significant improvement for those detected as at risk, benefits could outweigh the costs. Failure to meet this base rate criteria does, however, suggest the need to consider the implications of decisions made on the basis of screening procedures.

Selection Ratio. Just as the determination of base rate can be somewhat subjective and be influenced by district characteristics and criteria employed, selection ratio (which is defined as the percentage of students identified as at-risk) can also be arbitrary. Selection is often determined by failure to meet a previously determined cut-score on a screening test or battery as well as informal observations by teachers and other professionals. There is considerable variation in selection ratios reported in the literature. Additionally, the base rate and selection ratio employed in a particular study may differ markedly one from another. Differences between base rate and selection ratio percentages can have a significant effect on hit rate and predictive effectiveness as we shall observe.

Improvement Over Chance (IOC) Criterion. The majority of studies have been content to assess predictive effectiveness by reporting hit rates. The naive reader of such information has few anchor points to make sense out of a hit rate of,say, 85%. One might observe that 85% is less than perfect (100%) but much greater than 0% and better than 50%. Additional information can be had by asking how much better selection was obtained than could have been had by random selection. Improvement over chance prediction is similar to the concept of incremental validity (Sechrest, 1963) and

poses the question of how much the screening procedures contributed to the decision to which it was applied. The calculations employed take into account base rate, selection ratio, and test validity.

Validitycoefficients can also be calculated from prediction-performance matrices using the phi coefficient (Wiggins,1973). The computational formula is relatively simple:

$$Phi = \frac{P(VP) - BR * SR}{SQR\ ((BR)(1-BR)(SR)(1-SR))}$$

Where P(VP) = the % of valid positives among the total population
BR = base rate
SR = selection ratio

When the value of phi has been obtained, it is possible to estimate the probability of a valid positive in a new sample with a given base rate and selection ratio. The computational formula is:

$$P(VP) = BR * SR + phi\ (SQR((BR)(1-BR)(SR)(1-SR)))$$

(From Wiggins,1973)

This formula can be used to estimate the improvement in prediction with the addition of a test and can also be used to compare hit rates obtained with a screening test or procedure with those obtainable by chance. The results obtained by Reichmuth (1983) will be used to compare obtained results to chance predictions. (See Table 2)

Insert Table 2 About Here

Section a) of Table 2 shows the results obtained by Reichmuth. The overall hit rate (valid positives + valid negatives) was 80 percent. The notion of incremental validity can be applied to this study by asking how many correct decisions could be made by chance given a base rate of .15 and a selection ratio of .185. The formula for estimating P(VP) can be used for this purpose. Under random assignment, phi is equal to 0 and P(VP) = BR * SR = (.185)(.15) = .028. When the percentage of valid positives has been computed, the remainder of the matrix can be completed by subtraction. The outcomes of random assignment are presented in section b) of Table 2 and demonstrate that the overall hit rate under chance conditions would be (.028) + (.693) = .721. Thus, the use of the screening procedure in this study resulted in approximately 8 percent more correct decisions than would have occurred from random selection. Improvement over chance prediction (IOC) provides useful information in evaluating a screening procedure and some researchers have recently used this criteria.

Relative Improvement Over Chance (RIOC). It is important to note that in a great many studies of hit rate in the early identification literature, the base rate and selection ratio employed are rarely identical and are often quite discrepant. This discrepancy, as noted above, has a significant effect on the number of correct decisions possible. For example, if a study employs a base rate of 10 per cent and a selection ratio of 30 per cent, the maximum possible percentages of valid positives and valid negatives cannot exceed 80 percent. As Loeber and Dishion (1983) have noted in their review of studies predicting delinquency, percent improvement over chance (IOC) always falls between the chance value and the maximum correct value. Since studies vary with respect to maximum correct possible due to discrepancies between base rate and selection ratio, comparisons across studies are difficult. Loeber and Dishion propose an index called Relative Improvement Over Chance (RIOC) to provide a better basis for comparison. The formula for this index is: RIOC = (%IOC / %MC - %RC) x 100 where MC is the maximum correct and RC is random correct percentage. Data provided in the Loeber and Dishion review suggest that the RIOC index would be also useful for comparing studies in the early identification literature.

Expectancy Tables. While the prediction-performance matrix has obvious utility for research in evaluating screening programs, researchers also are responsible for communicating with teachers, administrators, and the public. While researchers can easily be entertained by hit rates, IOC's, and RIOC's, these concepts have been known to produce confusion, frustration, and sometimes drowsiness among those outside the research community. Better commmunication and

clarity can often be obtained by reporting screening results using expectancy tables (Wilson & Hewett, 1977). Such a format can also be helpful in clarifying the utility of a screening procedure. McCormick and Ilgen (1980) provide useful samples of both individual and institutional expectancy charts that are useful in presenting screening information. Table 3 presents a sample of such a presentation.

Insert Table 3 About Here

Sufficient Accuracy. The question posed in the title of the question asks when predictive accuracy is sufficient. While considerations of base rate, selection ratio, and test validity are necessary to answer such a question, they cannot be viewed as sufficient. The accuracy of the sorting process has to be viewed in the context of intervention decisions. When one is faced with the prospect of consuming a bad melon, the decision is typically to toss the melon and try again. We can't, however, toss students and we typically attempt to provide experiences for at-risk students that promote better learning. Ebel (1979) has suggested that our continuing focus on the problem of predictive validity is like raising a stir in the dust and then complaining that we can't see. The search for perfect predictive test validity has indeed resulted in considerable dust stirring and complaints that our predictive forsight is less than 20/20. Ebel has reminded us, however, that test scores and procedures are best used to encourage success rather than predict it.

In the context of early identification, we must ask how much our selection decisions contributed to the educational success of the children involved. The Reichmuth study yielded hit rates that could be viewed as respectable in comparison to similar studies in the literature. There was obviously some room for improvement, but the addition of better screening procedures would still result in a hit rate better than chance but less than perfect. The key question is the question of what happened to the students identified as at-risk. In the Reichmuth study, twenty-one students were so identified and parents of these children were advised to delay the child's entry into kindergarten. In this instance, 13 of the parents did so, but 8 of the parents decided to have their child enter kindergarten on schedule. Of these 8 who were in the second grade at the time of follow--up, 5 were reading above the 50th percentile (compared to other second graders). The thirteen who delayed were in the first grade at the time of follow-up and 8 of the 13 were reading above the 50th percentile (compared to other first graders)

and five were below that criterion. While the numbers are small, when these results are translated into simple expectancies, "high-risk" children who delayed a year had a better than even chance of being above average first grade readers. "High-risk" students who did not delay entry had a better than even chance of being above average second grade readers. These results suggest that the decision strategy to delay entry into kindergarten may not have been the best decision or method of intervention and suggested a need to consider alternatives in intervention.

In summary, predictive accuracy can and should be determined and appropriate use made of criterion to evaluate this accuracy. However, predictive accuracy may well approach perfection but such accuracy is sufficient only when students are provided better educational opportunity.

Table 1

Prediction-performance matrix for Reichmuth study

	low achievement	high achievement	
"at risk"	8 true positive	13 false positive	21
"ready"	9 false negative	83 true negative	96
	17	96	113 (Total N)

Table 2

Comparison of obtained percentages (Section a) with percentages expected by random assignment (Section b) given BR = .15 and SR = .185

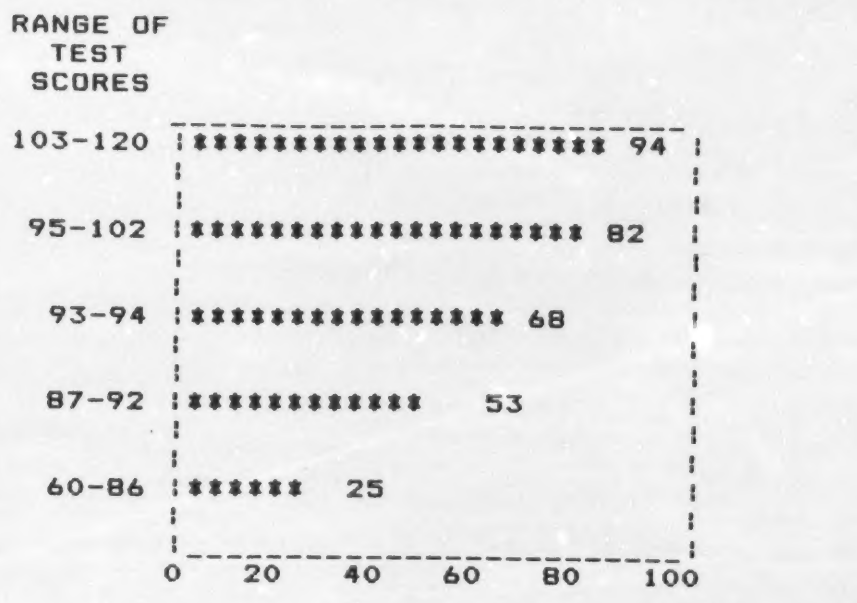
Section a) obtained results

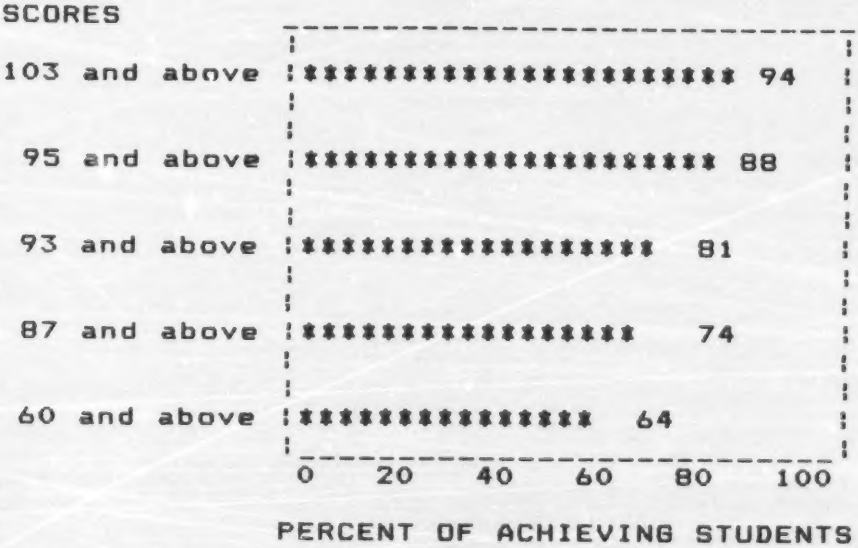
	low achievement	high achievement	
"at risk"	.07 true positive	.115 false positive	(SR = .185)
"ready"	.08 false negative	.735 true negative	(1 - SR = .815)
	17	96	
	(BR = .15)	(1 - BR = .85)	

Section b) random assignment

	low achievement	high achievement	
"at risk"	.028 true positive	.157 false positive	(SR = .185)
"ready"	.122 false negative	.693 true negative	(1 - SR = .815)
	(BR = .15)	(1 - BR = .85)	

Table 3
Illustration of (a) an individual expectancy chart and (b) an institutional expectancy chart as a function of scores on a screening battery for a sample of pre-school students. (Adapted from McCormick & Ilgen, 1980)





BIBLIOGRAPHY

Anastasi, A. (1982) Psychological Testing, Macmillan: New York

Barnes, K.E. Preschool Screening. Springfield: Charles Thomas.

Ebel, R.L. (1979) Essentials of Educational Measurement, Englewood Cliffs: Prentice-Hall

Eaves, L.C., Kendall,D.C., & Crichton, J.U. (1974). The early identification of learning disabilities: A follow-up study. Journal of Leaning Disabilities,7,632-638

Ferinden, W.E., Jacobsen,S., & Linden, N.J. (1970). Early identification of learning disabilities. Journal of Learning Disabilities,3,48-52

Lorber, R & Dishion, T. (1983) Early predictors of male delinquency: A review. Psychological Bulletin, 94,68-99

Keogh, B.K. (1977). Early ID: selective perception or perceptive selection? Academic Therapy,12,267-275

Keogh, B.K. & Becker, L.D. (1973) Early detection of learning problems: questions, cautions, and guidelines. Exceptional Children,40,5-11

Lichenstein, R. (1982) New instrument,old problem for early identification. Exceptional Children,49,70-72

Meehl, P.E. & Rosen,A. (1955) Antecedant probability and the efficiency of psychometric signs,patterns, or cutting scores. Psychological Bulletin, 52,194-216

Mercer, C.D. Students With Learning Disabilities. Columbus: Charles E. Merrill

Mercer, C.D., Algozzine,B.,& Trifiletti,J. (1979). Early identification - an analysis of the research. Learning Disabilities Quarterly, 2,13-24

Reichmuth, M. (1983). Evaluation of the predictive ability of four screening measures used in prekindergarten screening: A follow-up study. Unpublished specialist thesis, University of Northern Iowa, Cedar Falls, IA.

Satz,P., & Fletcher, J.M. (1979) Early screening tests: Some uses and abuses. Journal of Learning Disabilities, 12, 65-69

Taylor, H.C. & Russell, J.T. (1939) The relationship of validitycoefficients to the practical effectiveness of tests inselection. Discussion and tables. Journal of Applied Psychology, 23,565-578

Wilson, B. & Hewett, G. (1977, April). The identification of educationally handicapped children using existing information. Paper presented at the meeting of the American Educational Research Association, New York, NY

Wiggins, J.S. (1973) Personality and Prediction: Principles of Personality Assessment, Reading: Addison-Wesley